

NATALIA PAMUŁA-CIEŚLAK

Uniwersytet Mikołaja Kopernika w Toruniu

Instytut Informacji Naukowej i Bibliologii

Zjawisko Ukrytego Internetu – rola bibliotek w upowszechnianiu jego zasobów

Zjawisko Ukrytego Internetu (ang. Invisible Web) stanowi w dzisiejszym świecie duży problem i barierę dla użytkowników sieci WWW w dostępie do informacji, zwłaszcza sprawdzonej, recenzowanej, o kontrolowanej jakości i charakterze naukowym. Referat przedstawia to zjawisko, wskazując jednocześnie na ogromną rolę nowoczesnej biblioteki naukowej w upowszechnianiu wiedzy o takich zasobach i edukowaniu użytkowników na tym właśnie polu. Prezentuje także strategie wyszukiwania źródeł Ukrytego Internetu i charakteryzuje wybrane narzędzia, mogące być bibliotekom niezwykle pomocne w kształceniu i zaspakajaniu rosnących potrzeb informacyjnych swoich użytkowników. Stawia tezę, że taka działalność jest jednym z głównych zadań, stojących przed nowoczesnymi bibliotekami naukowymi XXI wieku.

1. Zjawisko Ukrytego Internetu

Ukryty Internet (ang. Invisible Web) nie jest zjawiskiem nowym. Już w roku 1994, a więc tuż po powstaniu sieci WWW, termin ten został użyty po raz pierwszy.¹ Zjawiskiem tym określa się zasoby elektroniczne (nierazko o bardzo wysokiej jakości informacyjnej) dostępne poprzez WWW, których wyszukiwarki internetowe nie mogą, bądź nie chcą, dodawać do swoich indeksów,² przez co zasoby takie stają się niewyszukiwalne w tradycyjny, popularny sposób. Z danych opublikowanych przez firmę BrightPlanet, zajmującą się badaniem Ukrytego Internetu, wynika iż wielkość tych zasobów jest ok. 500 raz większa, niż zasobów zgromadzonych na zaindeksowanych przez wyszukiwarki stronach WWW.³

2. Zawartość Ukrytego Internetu

Zawartość tematyczna Ukrytego Internetu kształtuje się następująco:⁴

- Nauki humanistyczne – 13,5%
- Wiadomości, media – 12,2%
- Informatyka – 6,9%

¹ A. Łamek: *Ukryty Internet*, Magazyn Internet 2002 nr 7, s. 58

² Ch. Sherman, G. Price: *The Invisible Web. Uncovering Information Sources Search Engines Can't See*, Medford, New Jersey 2003, s. 57

³ M. K. Bergman: *The Deep Web: surfacing hidden value*, The Journal of Electronic Publishing [online] Vol. 7 issue 1 [dostęp 01.07.2006]. Dostępny w Internecie: <http://www.press.umich.edu/jep/07-01/bergman.html>

⁴ M. K. Bergman: *op. cit.*

- Sztuki piękne – 6,6%
- Biznes – 5,9%
- Zdrowie – 5,5%
- Ludzie – 4,9%
- Edukacja – 4,3%
- Praca – 4,1%
- Nauki ścisłe – 4%
- Styl życia – 4%
- Prawo, polityka – 3,9%
- Informacje ze sfery rządowej – 3,9%
- Rekreacja i sport – 3,5%
- Podróże – 3,4%
- Zakupy – 3,2%
- Technika – 3,1%
- Rolnictwo – 2,7%

Nie bez znaczenia jest, iż ogromna część Ukrytego Internetu to dokumenty recenzowane lub takie, które zostały przygotowane pod kontrolą merytoryczną, literacką, językową i bibliograficzną. Do jego zasobów należą m. in.:

- Pliki nietekstowe (grafika, multimedia, dźwięki) – powodem jest to, że tego typu zasoby nie zawierają tekstu. Wyszukiwarki nie posiadają wzorca ani słów kluczowych (poza nazwą pliku – często nieznaczącą) do zaindeksowania takich zasobów. Rozwiązaniem mogły by być metadane opisujące obiekty nietekstowe;
- Bazy danych – problem z ich indeksowaniem leży w budowie baz. Roboty nie są w stanie zaindeksować, odwzorować i poprawnie wyświetlić informacji zorganizowanych w postaci pól i rekordów powiązanych ze sobą relacjami;⁵
- Serwisy płatnej rejestracji – roboty nie są dopuszczane do indeksowania takich zasobów ze względów komercyjnych;
- Serwisy bezpłatnej rejestracji – roboty nie są w stanie przejść formularzy rejestracyjnych, subskrypcyjnych i innych, wymagających reakcji użytkownika na elementy strony WWW.⁶

3. Rola bibliotek w upowszechnianiu zasobów Ukrytego Internetu

Obok istnienia Ukrytego Internetu można przejść obojętnie stwierdzając, że użytkownik poradzi sobie lepiej lub gorzej. Jednak w gąszczu wciąż przyrastającej ilości informacji w Internecie użytkownicy przestają sobie radzić z jej wyszukiwaniem. Stosują bowiem intuicyjnie jedną z najpopularniejszych i najmniej efektywnych strategii wyszukiwawczych. Polega ona na korzystaniu z jednej, wybranej wyszukiwarki internetowej i wpisywaniu do niej prostych zapytań. Jest to strategia zwana On the Web⁷ i sprawdza

⁵ Ch. Sherman, G. Price: *op. cit.*, s. 75

⁶ Ch. Sherman, G. Price: *op. cit.*, s. 73

⁷ Ch. Sherman, G. Price: *op. cit.*, s. 60

się głównie w przypadku haseł słownikowo-encyklopedycznych, rozrywki oraz popularnych serwisów WWW. W odpowiedzi użytkownik otrzymuje od wyszukiwarki miliony stron internetowych. Sprawdza kilka pierwszych rezultatów i jeśli tam nie znajduje tego, czego szuka – jest bezradny. Dlatego ważną rolę edukacyjną nowoczesnej biblioteki jest szkolenie użytkownika – tego realnego i wirtualnego – nie tylko w zakresie umiejętności wyszukiwawczych, ale i zastosowaniu właściwych narzędzi i strategii wyszukiwawczych w odniesieniu do zasobów Ukrytego Internetu.

4. Rola witryny internetowej biblioteki w upowszechnianiu zasobów Ukrytego Internetu

Warto propagować wśród użytkowników bibliotek, będących jednocześnie internautami, odmienną od strategii On the Web strategię Via the Web. Jest to jeden ze sposobów na Ukryty Internet. Polega na korzystaniu z różnych narzędzi wyszukiwawczych dostosowanych do potrzeb informacyjnych użytkowników bibliotek.⁸ Strategia ta pozwala na „szukanie głębiej” niż proponują to tradycyjne wyszukiwarki internetowe. Zakłada poszukiwania wielowarstwowe, wieloetapowe. Szczególnie dobrze sprawdza się przy wyszukiwaniu zasobów naukowych i badawczych, lecz z powodzeniem może być też stosowana do poszukiwań komercyjnych i informacyjnych. Aby przekonać użytkownika do takiej strategii wyszukiwawczej należy dostarczyć mu odpowiednich narzędzi. W przeciwnym razie zwróci się on w kierunku dotychczas stosowanej wyszukiwarki i jego poszukiwania będą nadal mało skuteczne.

Do promocji niestandardowych narzędzi i serwisów wyszukiwawczych doskonale nadaje się witryna internetowa biblioteki. Bardzo często na witrynach bibliotek znajdują się odnośniki do wielu ciekawych i pomocnych narzędzi wyszukiwania informacji, jednak bez wskazówek „po co”, „kiedy” i „jak” z nich korzystać użytkownik może mieć z tym spore problemy.

Atrakcyjną formą instruktazu wyszukiwawczego są szkolenia – kursy zamieszczane na witrynach bibliotek, gdzie w formie przystępnie napisanych instrukcji, z wplecionymi pomiędzy nie odsyłaczami do narzędzi i serwisów wyszukiwawczych, szkoli się użytkownika gdzie i jak szukać potrzebnych mu informacji w sposób szybki i efektywny.⁹

Oprócz kursów, z których użytkownik biblioteki, zwłaszcza ten wirtualny, będzie mógł skorzystać w dogodnym dla siebie momencie, te same narzędzia wyszukiwawcze warto zaprezentować w postaci odnośników podzielonych na działy, grupujące narzędzia ze względu na ich zastosowanie. Wówczas użytkownik, który zapoznał się z materiałami szkoleniowymi, będzie mógł korzystać z wybranych narzędzi bez poszukiwania ich na stronach kursu.

⁸ Ch. Sherman, G. Price: *op. cit.*, s. 60

⁹ Tego typu kursy zamieszcza na swoich stronach wiele bibliotek naukowych. Można je znaleźć na stronach wielu bibliotek szkół wyższych w Stanach Zjednoczonych [dostęp 01.07.2006], np. <http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/InvisibleWeb.html>, <http://www.library.albany.edu/Internet/deepweb.html>, <http://www.lib.lsu.edu/instruction/deepweb/deepweb-print.html>

5. Narzędzia wyszukiwawcze

W Internecie istnieje mnogość różnych narzędzi wyszukiwawczych poczynając od popularnych wyszukiwarek internetowych, poprzez katalogi tematyczne do narzędzi wyspecjalizowanych. Użytkownik może potrzebować każdego z tych narzędzi, aby je wykorzystać do różnych poszukiwań – musi tylko wiedzieć, że są i jak z nich właściwie korzystać.

Adresy wielu narzędzi wyszukiwawczych, podzielonych ze względu na ich zastosowanie, można odnaleźć na stronach serwisu internetowego poświęconego wyszukiwarkom i wyszukiwaniu – SearchEngineWatch.¹⁰ Zastosowano tu następujący podział narzędzi wyszukiwawczych:

1. Główne wyszukiwarki i katalogi internetowe;
2. Multiwyszukiwarki;
3. Wyszukiwarki wiadomości (newsów);
4. Wyszukiwarki, z którymi można zarobić (ang. Pay Per Click Search Engines)
5. Wyszukiwarki porównujące ceny i ułatwiające zakupy internetowe;
6. Wyszukiwarki multimedialne;
7. Narzędzia oferujące paski wyszukiwawcze (toolbars) oraz inne udogodnienia;
8. Wyszukiwarki dziecięce;
9. Wyszukiwarki krajowe – specjalizujące się w zasobach specyficznych językowo i w domenach narodowych danego kraju (inaczej zwane lokalnymi);
10. Wyszukiwarki specjalne – tu zastosowano dalszy podział na:
 - Wyszukiwarki udzielające odpowiedzi na pytania w języku naturalnym (ang. answers searching);
 - Wyszukiwarki komputerowe – wyszukujące materiały dotyczące informatyki i komputerów;
 - Wyszukiwarki domen;
 - Wyszukiwarki rządowe (przeszukujące oficjalne serwisy rządowe);
 - Wyszukiwarki Ukrytego Internetu;
 - Wyszukiwarki prawnicze;
 - Wyszukiwarki list mailingowych;
 - Wyszukiwarki medyczne;
 - Wyszukiwarki list dyskusyjnych, Usenetu, forów internetowych;
 - Wyszukiwarki naukowe;
 - Wyszukiwarki sklepowe;
 - Wyszukiwarki informacji turystycznej;
 - Wyszukiwarki zasobów WAP;
 - Inne serwisy specjalistyczne.

Jest to jeden z możliwych podziałów narzędzi wyszukiwawczych, który można i należy dostosowywać do potrzeb i wymagań użytkowników biblioteki i specyfiki polskiego

¹⁰ *Search Engine Watch* [online] [dostęp 01.07.2006]. Dostępny w Internecie: <http://www.searchenginewatch.com/links>

Internetu. Zabrakło tu chociażby np. wyszukiwarek branżowych – typu Yellow Pages oraz wyszukiwarek tematycznych (np. mediowych, specjalnych, wyszukiwarek zawartości baz danych).

Ważne jest uwzględnienie nie tylko narzędzi wyszukiwawczych, ale także serwisów informacyjnych. Za cenne należy uznać serwisy typu Subject Gateways oraz katalogi naukowe. Są one dziedzinowymi przewodnikami po zasobach internetowych, uporządkowanymi wg dziedzin. Są to zasoby wyselekcjonowane, oceniane, opisywane i katalogowane przez bibliotekarzy lub ekspertów dziedzinowych.¹¹ Wykaz tego typu serwisów znaleźć można w Biuletynie EBIB.¹²

Warto też zwrócić uwagę użytkowników na serwis poświęcony wyszukiwaniu informacji, kładący szczególny nacisk na zasoby Ukrytego Internetu. Jest to serwis Direct Search,¹³ prowadzony przez Gary Price'a – bibliotekarza amerykańskiego zajmującego się porządkowaniem Internetu, wyszukiwaniem informacji oraz zjawiskiem Ukrytego Internetu. Direct Search to stale rosnący zbiór odsyłaczy do serwisów i narzędzi wyszukiwawczych, ułatwiających dostęp do zasobów „ukrytych”. Na witrynie serwisu znajduje się rozbudowany interfejs wyszukiwawczy oraz uporządkowany zbiór opisanych i sprawdzonych zasobów elektronicznych, zwłaszcza tych przydatnych, które ciężko wyszukać w sposób tradycyjny.

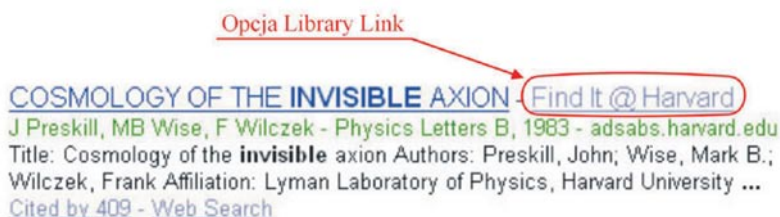
Szczególne miejsce wśród specjalnych i wyspecjalizowanych narzędzi i serwisów wyszukiwawczych powinny zajmować wyszukiwarki naukowe. W nieskomplikowany, znany użytkownikom tradycyjnych wyszukiwarek, sposób można w nich wyszukać informacje z dziedziny nauki, badań i edukacji. Zasoby, które przeszukują tego typu narzędzia to czasopisma recenzowane dostępne on-line, pełne teksty oraz abstrakty artykułów z nich pochodzących, bazy danych gromadzące informację edukacyjną i naukową, katalogi biblioteczne i bibliografie, roczniki statystyczne, serwisy pełnotekstowe, e-booki i e-printy, materiały konferencyjne, a także źródła nietekstowe: mapy i systemy nawigacyjne, rysunki techniczne, dane numeryczne i teleadresowe. Za sprawą umów z wydawcami roboty takich wyszukiwarek mają także dostęp do płatnych czasopism i baz danych (będących istotną częścią Ukrytego Internetu). W przypadku zasobów płatnych użytkownik wyszukiwarki może zobaczyć jedynie opis bibliograficzny i abstrakt tekstu, zaś pełny tekst jest dostępny po dokonaniu rejestracji w systemie (opcja dla subskrybentów). Jest to jednak cenna informacja dla użytkownika – uzyskuje wiedzę, gdzie znajdują się dane, których potrzebuje. Tu również dużą rolę powinny odegrać nowoczesne biblioteki naukowe, będące subskrybentami wielu serwisów płatnych. Mogą one uświadomić użytkownikowi, że za pośrednictwem biblioteki ma on dostęp do potrzebnych mu zasobów płatnych. W ten sposób wiele serwisów komercyjnych, będących w istocie częścią Ukry-

¹¹ L. Derfert-Wolf: *Serwisy tematyczne o kontrolowanej jakości w Internecie – subject gateways*, Biuletyn EBIB [online] 2004 nr 6 [dostęp 01.07.2006]. Dostępny w Internecie: <http://ebib.oss.wroc.pl/2004/57/derfert.php>

¹² L. Derfert-Wolf, *Serwisy tematyczne o kontrolowanej jakości w Internecie – subject gateways. Wykaz*, Biuletyn EBIB [online] 2004 nr 6 [dostęp 01.07.2006]. Dostępny w Internecie: <http://ebib.oss.wroc.pl/2004/57/wykaz.php>

¹³ G. Price: *Direct Search* [online] [dostęp 01.07.2006]. Dostępny w Internecie: <http://www.freepint.com/gary/direct.htm> [dostęp 01.07.2006]

tego Internetu – tą najwartościowszą i najbardziej „ukrytą” – dla sporej części użytkowników „ukrytymi” być przestanie. Gotowym narzędziem służy wyszukiwarka naukowa Google Scholar,¹⁴ która będąc częścią koncernu Google, zdobyła już popularność wśród internautów. Pod adresem internetowym <http://scholar.google.com/intl/en/scholar/libraries.html> znajdują się istotne informacje o tym, w jaki sposób biblioteka może przyłączyć swe zasoby do Google Scholar i zaprezentować je użytkownikom w rankingu wyszukiwawczym. Może to być istotną częścią marketingu bibliotecznego i promocji własnych zbiorów – nie tylko elektronicznych. Internauta, który zaloguje się w serwisie wyszukiwawczym jako użytkownik biblioteki, która zarejestrowała swe zbiory w wyszukiwarce, w wynikach wyszukiwania będzie mógł zobaczyć, do których z dokumentów ma dostęp pełnotekstowy dzięki swojej bibliotece.



Rys. 1 Google Scholar – rezultat wyszukiwania z zaznaczoną nazwą biblioteki posiadającej pełny tekst wyszukanego artykułu

6. Zakończenie

Czy biblioteki XXI wieku przetrwają? Wydaje mi się, że trudno by zniknęły. Jednak dużo ważniejszym pytaniem wydaje się być – w jakiej formie przetrwają? Czy będą niczym stare muzea gromadzące tysiące eksponatów, do których ktoś sięga z rzadka, czy przetrwają jako centra łączące w sobie tradycję z nowoczesnością? Sądzę, że dla bibliotek, a przede wszystkim ich użytkowników wizja druga jest znacznie atrakcyjniejsza. Biblioteka powinna stać się nie tylko budynkiem, ale i wirtualną przestrzenią, w której użytkownik zawsze i w prosty sposób znajdzie informację, której potrzebuje. Do tego niezbędni będą nowocześni bibliotekarze, zdolni nie tylko udzielić bezpośredniej informacji na pytanie realnego użytkownika biblioteki, ale umiejący też dostarczyć mu tę wiedzę wirtualnie – za pomocą nowoczesnej witryny internetowej biblioteki. Bibliotekarze wykorzystujący swą wiedzę o narzędziach i serwisach wyszukiwawczych, o istnieniu Ukrytego Internetu będą niezbędni w XXI wieku – wieku informacji, społeczeństw informacyjnych i „googlizacji” – by uporządkować chaos informacyjny w sieci WWW i taki uporządkowany Internet prezentować użytkownikom. Mam nadzieję, że użytkownicy to docenią.

¹⁴ Google Scholar [online] [dostęp 01.07.2006]. Dostępny w Internecie: <http://scholar.google.com>

Bibliografia

1. Bergman M. K.: *The Deep Web: surfacing hidden value*, The Journal of Electronic Publishing [online] Vol. 7 issue 1 [dostęp 01.07.2006]. Dostępny w Internecie: <http://www.press.umich.edu/jep/07-01/bergman.html>
2. Derfert-Wolf L.: *Serwisy tematyczne o kontrolowanej jakości w Internecie – subject gateways*, Biuletyn EBIB [online] 2004 nr 6 [dostęp 01.07.2006]. Dostępny w Internecie: <http://ebib.oss.wroc.pl/2004/57/derfert.php>
3. Derfert-Wolf L.: *Serwisy tematyczne o kontrolowanej jakości w Internecie – subject gateways. Wykaz*, Biuletyn EBIB [online] 2004 nr 6 [dostęp 01.07.2006]. Dostępny w Internecie: <http://ebib.oss.wroc.pl/2004/57/wykaz.php>
4. Łamek A.: *Ukryty Internet*, Magazyn Internet 2002 nr 7, s. 58-60
5. Sherman Ch., Price G.: *The Invisible Web. Uncovering Information Sources Search Engines Can't See*. Medford, New Jersey 2003

Wykaz źródeł elektronicznych wykorzystanych w publikacji

1. *Google Scholar* [online] [dostęp 01.07.2006]. Dostępny w Internecie: <http://scholar.google.com>
2. *Google Scholar Support for Libraries* [online] [dostęp 01.07.2006]. Dostępny w Internecie: <http://scholar.google.com/intl/en/scholar/libraries.html>
3. *Invisible or Deep Web: What it is, Why it exists, How to find it, and Its inherent ambiguity* [online] [dostęp 01.07.2006]. Dostępny w Internecie: <http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/InvisibleWeb.html>
4. *LSU Libraries, The Deep Web* [online] [dostęp 01.07.2006]. Dostępny w Internecie: <http://www.lib.lsu.edu/instruction/deepweb/deepweb-print.html>
5. Price G.: *Direct Search* [online] [dostęp 01.07.2006]. Dostępny w Internecie: <http://www.freepint.com/gary/direct.htm>
6. *Search Engine Watch: Tips About Internet Search Engines & Search Engine Tips* [online] [dostęp 01.07.2006]. Dostępny w Internecie: <http://www.searchenginewatch.com>
7. *Search Links: Search Engine Worldwide* [online] [dostęp 01.07.2006]. Dostępny w Internecie: <http://www.searchenginewatch.com/links>
8. *The Deep Web* [online] [dostęp 01.07.2006]. Dostępny w Internecie: <http://library.albany.edu/Internet/deepweb.html>

The Invisible Web – a role of libraries to widespread that resources

The Invisible Web is a big problem in today's world. especially in access to reviewed, high quality, scholarly and scientific information sources. This paper presents that problem and show a hudge role of modern academic library to dissaminate a knowledge about that kind of sources and educate users in that field. It also presents a searching strategies for Invisible Web sources and discribe a selected tools, which could be able for libraries to educate and satisfy a growing information needs of they users. It says that this information activity is one of the main task of modern academic libraries of 21 century.